Content Validity of Manual Spinal Palpatory Exams. A Systematic Review.

Wadie I. Najm, MD

Associate Clinical Professor

University of California, Irvine

Pomona, CA 91766-1854

smishra@earthlink.net

Vivian Dickerson, MD

Associate Clinical Professor

University of California, Irvine

101 The City Drive

Orange, CA 92868

vdickerson@socal.rr.com

Alan Adams, DC, MS

Faculty Administrator,

Florida State University

110J-Westcott Building

Tallahassee, FL 32306

aadams@mailer.fsu.edu

Sibylle Reinsch, PhD

Research Associate, Dept. of PM&R

University of California, Irvine

101 The City Drive

Orange, CA 92868

sreinsch@uci.edu

Linda L. Suk-Ling MURPHY, MLS

Health Sciences Librarian

University of California, Irvine

231 Science Library

Irvine, CA 92697

lmurphy@uci.edu

Arnold F. Goodman, Ph.D.

Associate Director Center for Statistical Consulting

4900 Berkeley Place

Irvine, CA 92697

agoodman@uci.edu

Address correspondence to:

Wadie I. Najm, MD

University of California, Irvine

101 The City Drive Bldg. 200 #512, Rt. 81

Orange, CA 92868

Abstract

Background: Many health care professionals use spinal palpatory exams as a primary and well-accepted part of the evaluation of spinal pathology. However, few studies have explored the validity of spinal palpatory exams. To evaluate the status of the current scientific evidence, we conducted a systematic review to assess the content validity of spinal palpatory tests used to identify spinal neuro-musculoskeletal dysfunction. **Methods:** Review of eleven databases, and a hand search of peer-reviewed literature were undertaken from 1965-2001. Five papers met the inclusion/exclusion criteria. Two blinded reviewers abstracted pertinent data from the five papers, using a specially developed quality-scoring instrument. **Results**: Two papers focused on identifying the level of fixation and one focused on range of motion. All three studies used a mechanical model as a reference standard. Two papers explored the validity of pain assessment using the visual analogue scale or the subjects' own report as reference standards. Overall the sensitivity of studies looking at range of motion tests and pain varied greatly. Poor sensitivity was reported for range of motion studies regardless of the examiner's experience. A slightly b 1 Tf 6.186(h)-4(t)-2r

Background:

Injury of the spine and back are classified as the most frequent cause of limited activity among people younger than 45 years[1-2]. Approximately 10 percent of the adult population has neck pain at any one time[3], and 80% of the population will experience low back pain (LBP) at some time in their lives[4]. Five to ten percent of the workforce is off work annually because of LBP. Indeed, LBP is second only to headache among the leading causes of pain. Approximately 80-90% of LBP is mechanical (non-organic musculoskeletal dysfunction) in origin[5]. Patients with mechanical spinal pain often seek and receive spinal manipulation by chiropractic, osteopathic and allopathic clinicians, physical therapists or other health care professionals[6].

Health care professionals have utilized spinal palpatory diagnostic procedures and manual manipulative treatment for several millennia to treat back injury and pain [7-8.] Along with the history of illness and physical exam, examiners utilize specific spinal palpatory diagnostic tests in order to identify spinal neuromusculoskeletal dysfunction. Spinal neuromusculoskeletal dysfunction refers to an alteration of spinal joint position, motion characteristics and/or related palpable paraspinal soft tissue changes. Outcomes and effectiveness of spinal manipulative procedures rely on appropriate and skilled treatment that is based on an accurate diagnosis, which in turn depends upon the accuracy of the palpatory procedures used.

Spinal palpatory procedures have been described in journals[9-11] and textbooks[12-19]. Static palpation of anatomical landmarks for symmetry, palpation of spinal vertebral

joints before, during and after active and passive motion tests, spinal and paraspinal soft tissue palpatory assessment for abnormalities or altered sensitivity are most common.

Several narrative reviews of the literature on the validity and reliability of spinal palpatory diagnostic procedures have been published[20-27]. However, most reviews are discipline-specific despite the fact that similar spinal palpatory procedures are used across disciplines. Only two systematic reviews of spinal palpatory validity studies have been published[28-29]. One study was a limited review of chiropractic literature on palpatory diagnostic procedures for the lumbar-pelvic spine[28] and the other concentrates on validity studies at the sacroiliac joint[29]. An annotated bibliography[30] and a systematic review of the primary reliability research studies published between 1971 and 2001 are in progress.

Validity and reliability are concepts that are often used interchangeably, but the concepts are quite different. Validity is the accuracy of a measurement of the true state of a phenomenon[31], while reliability measures the concordance, consistency or repeatability of outcomes[24]. However, even if a measurement is consistent and reliable, it is not necessarily valid (e.g., an arrow may consistently hit the target area, but never hit the bulls-eye).

There are various types of validity studies. The concept of validity differs in qualitative and quantitative research[31]. Though it can be argued that palpatory

diagnostic procedures are subjective and therefore qualitative, investigators in the field believe they can measure a physiological phenomenon that can be detected by objective means. They maintain that studies addressing the validity of spinal palpatory diagnostic tests are quantitative studies. The types of quantitative validity studies can be distinguished as follows: face validity; construct validity, criterion validity and content validity.

Face validity is the extent to which a test appears to measure what it is supposed to measure. In other words, whether the proposed test seems to provide a reasonable measure of the concept it is intended to measure. For example, spinal vertebral joint motion palpation tests, which aim to detect the presence of hypomobility, have face validity because they seem to be reasonable measures of the concept they are intended to measure[32]. Face validity studies have been criticized for being subjective, intuitive and unsubstantiated. Troyanovich and Harrison[32] pointed out that in spite of the common perception or belief that motion tests are valid and reliable for assessment of presence or absence of restricted vertebral motion, there was no evidence to support this concept. Thus, palpatory vertebral motion diagnostic tests are prime examples of tests accepted on face validity.

Construct validity is the extent to which a test identifies the concept or trait of that which is being measured. A construct is a hypothetical or conceptual idea that may be used to label or explain observed phenomenon[33]. For example, taking a dysfunctional vertebral joint as the concept, a test demonstrating the ability to identify the presence or

absence of that concept or its related components, is said to have construct validity. Feinstein describes construct validity as an appraisal of the effectiveness with which a measure does its job in describing an existing or established construct; i.e. does the measure behave the way one would predict on the basis of the concept it represents? For example, Jull et al[34] compared cervical spinal static palpation to diagnostic nerve blocks with anesthesia. The construct is that tenderness upon provocative palpation is related to local nerve irritation and nerve conductivity. A local anesthetic nerve block of related spinal segments showed that the identified tender spots no longer elicited a pain response. They thus demonstrated that there is a high degree of correlation between the palpatory test that identified a tender spot and the ability of the anesthesia to reverse the results of the provocative test. Therefore, the pain provocative palpatory tests used were demonstrated to have high construct validity.

Construct validity, however, is an artificial framework that is not directly observable[27]. To establish construct validity of a test or measure, the researcher must determine the extent to which the measure correlates with other measures designed to measure the same thing and whether the measure behaves as expected. Construct validity studies do not measure the same phenomena that palpatory procedures are designed to measure (i.e., resistance to digital pressure or motion), but similar pher(y)2(a)-4(r)3( )278]d0(e)-thasiasuretay outd teta( )-2(t)-3(e)-2(tc3 -2(h)-4(i)24(n)-4(ce)6( )(y)2(a)-4

There are other examples of construct validity studies using instruments to measure skin temperature, electrical skin resistance and/or gross range of motion to discern a dysfunctional vertebral segment. These measurements are then compared to those obtained by another examiner who utilizes one or several palpatory procedures that assess resistance to joint motion or paraspinal soft tissue abnormalities to help to discern a dysfunctional vertebral segment. Or one examiner uses pain provocation and the other palpatory motion restriction sense to assess for a dysfunctional vertebral segment.

Criterion validity measures the extent to which an intervention allows a researcher to predict behavioral or pathological outcomes. Criterion validity studies, therefore, do not measure the phenomena being palpated, but attempt to correlate the findings of a palpatory procedure (e.g.) with another measurable outcome like diagnosed visceral disease. For example, Beal[38]and Tarr[39] studied the ability of physicians using spinal palpatory procedures to identify, or predict, which patients had visceral disease related to the spinal findings of altered structure, motion and/or soft tissue.

Content validity is the extent to which a measure adequately,comprehensively

A reference standard (also called “gold standard”) is a measure accepted by consensus of content experts as the best available for determining the presence or absence of a particular phenomenon. When there is no perfect reference standard, as in the case of measurement of a patient’s sense of pain provocation, i.e., pressing on a “tender point” or “trigger point”, then pragmatic criteria can be used as a reference standard[41]. The visual analog pain scale has been used as a pragmatic reference standard for palpatory pain provocati

valid, we conducted a systematic review to assess the content validity of spinal palpatory tests used to identify spinal neuro-musculoskeletal dysfunction.

METHODS

Study Setting

The study was conducted at the Susan Samueli Center for Complementary and Alternative Medicine (University of California-Irvine [UCI]). A multi-disciplinary team of clinicians, researchers, a statistician, and a health sciences librarian participated in the systematic review. The clinicians represented content area expertise in osteopathic and chiropractic medicine, family medicine, and clinical research. In addition, the researchers had expertise and experience in evidence-based medicine, research design and methodology.

The study inclusion/ exclusion criteria were adapted and modified from those published previously by the Cochrane Collaboration[44] and others[45-46]. Studies included in the review met the following four criteria: 1) the studies pertained to manual spinal (cervical, thoracic, lumbar, and surrounding para-spinal soft tissue but not the sacrum or pelvis) palpation procedures; 2) the studies included measurement of validity or accuracy of spinal palpation, where validity was defined as the capability of the manual spinal palpation procedure to do what it is supposed to do and accuracy was defined as a measure of how well it actually does that; 3) the studies were dissertations or a primary research studies published in a peer-reviewed journal; 4) the document could be written in any language; 5) the primary research must have been published or

accepted for publication; and 6) all studies were made available between January 1, 1966 and September 30, 2002. Studies were excluded from the review based on the

search outcome of individual databases, to 11 databases that had a potential coverage for the areas of osteopathic medicine, allopathic medicine, chiropractic, and physical therapy. The databases accessed by the project included: PubMed MEDLINE, MANTIS, CINAHL, Web of Science, Current Contents, BIOSIS, EMBase OCLC FirstSearch, Cochrane, Osteopathic Database, and Index to Chiropractic Literature. The selection of databases was based primarily on the availability of online resources that we could access from our affiliated institution libraries.

In addition to the online literature search strategy, we used manual methods to identify appropriate literature. These manual methods included gleaning references that were cited in studies selected from the online search, consulting experts in the fields of chiropractic and osteopathic medicine, contacting authors of eligible conference abstracts, and manually searching bibliographies of osteopathic text-books and review articles on somatic dysfunction.

We used a three-step selection process to identify articles for the systematic review. First, we reviewed titles identified through the online search, and excluded those which gave no indication that the studies pertained to validity. Second, we reviewed the abstracts of all the remaining studies identified through the application of our search template, and excluded studies that did not meet the inclusion criteria. Third, we reviewed the complete paper and applied the inclusion/exclusion criteria to studies included at step two.

“relative” score (i.e., [absolute score/ total score that could be obtained] X 100). The relative score was especially important for studies wherein certain aspects of the quality scoring components were inapplicable (i.e., the subjects’ criteria was inapplicable for studies which used mechanical models or measures). An article’s score (absolute or relative) indicated its quality in terms of its internal validity criteria (whether conclusions drawn from study are likely to be unbiased) and the authors’ explicit description of the

The three motion palpation studies were done in the UK. All three studies utilized mechanical models as the study subjects as well as the reference standard. The two pain studies were done in Sweden. One study (Kristiansson), recruited only pregnant female subjects (n=200, representing a 90% response rate: 200/222), while the other study recruited an entirely male population (n=75, they failed to report the response rate)

Reliability of the palpation procedure was not reported in any papers with the exception of 1 (Jensen) looking at motion palpation in a mechanical model.

Quality Scoring Findings:

In general the quality score would indicate the rigor with which the science was presented in the paper. Quality scores of included studies ranged from 45.5 to 82 out of a possible100. The overall quality of the included studies was good for those focusing on motion palpation (69.5 - 82), and fair for those looking at pain (45.5 - 55.5). Discussion of examiners and study conditions were the two major areas where weakness was noted in the two pain studies, but not in the motion palpation studies. Statistical tests used were adequate for all studies (this was one of the inclusion criteria). All studies were done in the 1990's; hence the time factor was not felt to be contributive.

INSERT TABLE 4 ABOUT HERE

Study findings:

Motion Palpation Tests: The three studies examining motion palpation were similar in using a mechanical model as the reference standard and focusing on the lumbar spine only. While two studies used similar examiner groups and motion test, the third study (Moruzzi) looked only at one group of examiners using two different motion test procedures.

Two studies (Harvey & Jensen) looked at intersegmental motion restriction, using sagital and coronal motion as determined by two groups of chiropractic examiners with

different experience levels (senior students and practitioners). Both studies presented data on sensitivity (ability of a test to detect correctly restricted motion segments) and specificity (ability of a test to detect correctly unrestricted motion segments). The sensitivity for both groups in each study varied between 0.510 and 0.636, and the specificity from 0.868 to 0.902, indicating less ability to detect restricted motion segments than unrestricted motion segments. The sensitivity for practitioners in both studies was poor (0.478 and 0.526). For students, the sensitivity was lower in the Harvey study (0.538) than the Jensen study (0.72).

Based on the data provided in each of the studies we calculated the positive and negative predictive (PPV; NPV) values and the likelihood ratio (LR) for each group. The PPV was less than 50% in both studies, for both groups (42.3-46.2%) and for each subgroup. While the NPV was greater than 80% (878\4( )-2(8)(\4( )-2.9133 0 Td [(4)-4(6)-4(.)8(93(o)-4

Given that face, construct and criterion validity studies do not measure the phenomena being palpated, but attempt to correlate the findings of a palpatory

Unfortunately, most of the research study results reported are not comparable due to variability in the palpatory tests, terminology, research design, methodology and statistical analysis utilized. These inconsistencies make it difficult to rate the relative value of their results. There is a worldwide concerted effort underway to rectify this problem. The International Federation of Manual Medicine (FIMM), an international organization of physicians and surgeons who practice manual medicine held their General Assembly in Chicago in July 2001. At that meeting, their Scientific Committee reported that their top priority is to promote validity, reliability, sensitivity and specificity studies of spinal palpatory diagnostic procedures. They recently developed guidelines ("Protocol Formats") on how to perform high quality validity and reliability studies of spinal palpatory procedures, which are available on their web site http://www.fimm-online.org/. They recommend the use of valid palpatory tests so that homogeneous populations with spinal musculoskeletal dysfunction can be selected and treated as part of a controlled clinical trial. The results of these trials can subsequently be combined using meta-analysis and would help formulate guidelines for the practice of spinal manipulation.

It is difficult to translate these results into the clinical setting due to the limited number of studies, focused anatomical sites and populations studied. All three-motion palpation studies used a mechanical model as the subjects and reference standard, and focused

strategy allowed more clarity since only content validity studies, which attempt to measure the same phenomena as that which is being palpated, were included in this systematic review. Despite the number of safeguards used to be inclusive (multiple databases, hand search, review by experts, and multiple searches) in our search, a few studies published but not included in these databases could have been missed.

The quality assessment tool, used for this review, was developed by this team of researchers based on their evaluation of the literature, feedback from methodologists and statisticians. Although we feel that the instrument is well balanced and unbiased, it might have over or underestimated the quality of certain papers. When comparing the quality scores assigned to studies included in this paper to scores assigned to the same papers in other systematic reviews[27], one notes that our scores are consistently lower.

Conclusion:

Despite the use of manual palpatory spinal palpation by many health care disciplines, very few studies investigated it's ability to measure what it intends to measure (content validity). Given the high frequency of spinal pathology and the use of these diagnostic methods, well-designed studies are needed. For the practice of evidence-based medicine, it is important to assess the efficacy and effectiveness of procedures usually and customarily used in clinical practice. To this end, established benchmarks for the validity and reliability of procedures are essential.

This comprehensive systematic review has highlighted serious gaps in our knowledge about the accuracy of spinal palpatory procedures. The findings have

implications for research, clinical practice, and policy. From the research perspective, researchers across discipline need to incorporate more rigor towards the definition of the study questions, methods and measures, implementation procedures, and reporting. The absence of well identified reference standards and possible technical difficulties conducting these studies might have contributed to this scarcity.

From the clinical perspective, the findings suggest poor sensitivity of the range of motion and pain diagnostic tests in the evaluation of spinal dysfunction. From a policy perspective, given that manual procedures are a cornerstone towards diagnostic and therapeutic interventions across disciplines, professional societies and associations need to enact continuing medical education and research guidelines to address the efficacy of spinal palpatory procedures.

10. Walker BF and Buchbinder R. Most commonly used methods of detecting spinal subluxation and the preferred term for its description: A survey of chiropractors in Victoria, Australia. J Manipulative Physiol Ther, Nov/Dec, 1997; 20(9):583-589.
11. Walker BF. The reliability of Chiropractic methods used for the detection of spinal subluxation. An overview of the literature. ACO March 1996; 5(1):12-22.
12. Ward RC Ex Ed. Foundations for Osteopathic Medicine. Philadelphia: Lippincott, Williams and Wilkins. 2003.
13. Chaitow L. Palpation skills: Assessment and diagnosis through touch. New York: Churchill-Livingstone. 1997.
14. Bergman TF, Peterson DH and Lawrence DJ. Chiropractic techniques. Principles and procedures. New York: Churchill-Livingstone. 1993.
15. Kaltenborn FM. The spine: Basic evaluation and mobilization techniques. Minneapolis, MN: OPTP. 1993.
16. Maitland GO ed. Maitland's vertebral manipulation. Oxford: Butterworth-Heinemann. 2000.
17. Lewit K. Manipulations therapy in rehabilitation of the locomotor system. Boston: Butterworths. 1985.
18. Maigne R. Orthopedic medicine. A new approach to vertebral manipulation. WT Liberson, trans. Springfield, IL: Charles Thomas. 1972.
19. Wadsworth CT. Manual examination and treatment of the spine and extremities. Baltimore: Williams and Wilkins. 1988.

20. Johnston, W.L. Interexaminer reliability studies: spanning a gap in medical research. Louisa Burns Memorial Lecture. J Am Osteopath Assoc 1982, 81:819-829.
21. Gonnella C, Paris SV, and Kutner M: Reliability in evaluating passive intervertebral motion. Phys. Ther. 1982, 62:436-444.
22. Panzer, DM: The Reliability of Lumbar Motion Palpation. J Manipulative Physiol Ther. October 1992, 15(8): 518-524.
23. Keating JC, Jacobs GE. Inter- and intraexaminer reliability of palpation for sacroiliac joint dysfunction. JMPT 1989;12(2):155-8.
24. Haas M. The reliability of reliability. JMPT 1991;14(3):199-208.
25. Russell R. Diagnostic palpation of the spine: A review of procedures and assessment of their reliability. J Manipulative Physiol Ther 1983; 6:181-183.
26. Alley JR. The clinical value of motion palpation as a diagnostic tool: A review. The Jouu

30. Seffinger MA, Adams AA, Najm WI, Dickerson V, Reinsch S, Goodman A, Gokhale DV, Murphy L, Mishra S. Spinal Palpatory Diagnostic Procedures Utilized by Practitioners of Spinal Manipulation: Annotated Bibliography of Reliability Studies. (Accepted for publication, JCCA 2003.)
31. Winter G. A comparative discussion of the notion of 'validity' in qualitative and quantitative research. The Qualitative Report. Vol. 4, Numbers 3 & 4, March 2000 (http://www.nova.edu/ssss/QR/QR4-3/winter.html)
32. Troyanovich SJ and Harrison DD. Motion palpation: Its time to accept the evidence (y)(n) and comment. J Manipulative Physiol Ther, Oct 1998; 21J(Q)-4(8)(9)-4(8)6(;)-2( )-2(2)

39. Tarr RS, Feely RA, Richardson DL, Mulloy AL, Nelson KE, Perrin WE, Allin EF, Efrusy ME, Greenstein SI, Vatt RD. A controlled study of palpatory diagnostic procedures: assessment of sensitivity and specificity. J Am Osteopath Assoc. 1987 Apr;87(4):296-301.
40. Troyanovich SJ and Harrison DD. The reliability and validity of chiropractic assessment procedures. Chiropractic Technique, Feb. 1996; 8(1):10-13.
41. Knottnerus JA, ed. The evidence base of clinical diagnosis. London: BMJ books, 2002.
42. Haldeman S, Chapman-Smith D, Peterson DM Jr. eds. Guidelines for chiropractic quality assurance and practice parameters. Proceedings of the Mercy Center Consensus Conference; 1993 January 25-30. Burlingame, CA: Aspen Publishers; 1993.
43. Troyanovich SJ and Harrison DD. In reply to letter to the editor. J Manipulative Physiol Ther, March/April 1999; 22(3):182-183.
44.

studies: dida

55. Juni P, Wistchi A, Bloch R, Egger M. The Hazards of Scoring the Quality of Clinical trials for Meta-analysis. JAMA 1999;282(1):1054-1060.
56. Knottnerus, Andre. The Evidence Base of Clinical Diagnosis. BMJ Books, January 2002.
57. Olson SL, O'connor DP, Birmingham G, Broman P, Herrera L. Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. J Orthop Sports Phys Ther. Jan 2000; 30(1): 13-20
58. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analog scales as ratio scale measures for chronic and experimental pain. Pain. 1983; 17:45-56.

63. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001;285:1987-91.

Studies included in this review:

1. Harvey D, Byfield D. Preliminary Studies with a Mechanical Model for the Evaluation of Spinal Motion Palpation. Clinical Biomechanics 1991; 6:79-82
2. Jensen K, Gemmell H, Thiel H. Motion palpation accuracy using a mechanical spine model. European J Chiropractic 1993;41:67-73.
3. Kristiansson P, Svardsudd K, von Schoultz B. Back Pain During Pregnancy. A Prospective Study. Spine 1996;21:702-8.
4. Moruzzi S, Accuracy of Two Different Motion Palpation Procedures For determining Fixations in the Lumbar Spine Using An Articulated Spinal Model. Dissertation. Anglo-European College of Chiropractic, 1992-1993.
5.

TABLE 1

Studies reviewed and excluded from this study of content validity

Author	Validity of Spin Palpatory Procedures	Validity	Statistics used	Discipline	Palpatory method used	Reference test used
Beal 1989	No	Construct	Descriptive	DO	Passive ROM; tissue texture; passive mobility at specific spinal levels	Electromyography
Braun 1991	No	Construct	Pearson Correlation coefficient	PT	Cervical Range of Motion instrument	Questionnaire and Muscle Index neck subscale
Bush 2000	No	Construct	Correlation Coefficient	PT	Single, double and stabilization methods measuring Cervical ROM	X-Ray
Gregory 1998	Yes	Construct	Chi square	DC	Motion palpation	X-Ray
Haas 1994	No	Construct	fisher 2 tailed exact t test/ Prevalence	DC	Vertebral challenge	Response to manual high velocity low amplitude adjustment
Haas 1995	Yes	Construct	Descriptive	DC	Motion palpation using a spinous process contact (Bergmann)	Response to manual high velocity low amplitude rotatory manipulation
Johnston 1985	Yes	Construct	MANOVA	DO	Range of motion; passive and active ROM	Kinemetic procedure
Jull 1988	Yes	Construct	Sensitivity & Specificity	PT	Passive accessory intervertebral movements	Cervical nerve block
Kawchuk	No	Construct	Intraclass correlat	DC	Tissue compliance meter	Measurement on a controll

1995		coefficient	surface

Olson 2000	Yes	Predictive/ Criterion	Pearson Correlation coefficient	MD	Pressure Algometry for tender point sensitivity; ROM	Visual Analog Scale
Tarr 1987	Yes	Predictive/ Criterion				

TABLE 2

QUALITY SCORING CRITERIA, TOTAL WIEIGHT AND TOTAL SCORE ASSIGNED.

	CRITERIA	TOTAL WEIGHT	TOTAL SCORE
1	STUDY SUBJECTS • Study Subjects Adequately Described • Presentation Characteristics • Spectrum of severity of Symptoms • Subject Selection Criteria • Number of Subjects in Study detailed • Sample Size Determined by Power Analysis • Number Subjects Completed Study • Recruitment Procedure	15%	15
2	EXAMINERS • Selection Criteria for Examiners Described • Background of Examiners Described	10%	10
3	REFERENCE STANDARD • Was Reference Standard used • Reference standard procedure described / referenced • Expected Outcome Described • Validity of reference Standard • Reliability of Reference Standard • Positive or Negative Test Result Defined	15%	15
4	PALPATORY TEST • Description of palpatory test Procedure • Expected Outcome Described • Reliability of Test Described • Positive or Negative Test Result Defined	15%	15
5	STUDY CONDITIONS	15%	15

	• Time Interval for Test/ Retest procedure • Examiner blinded to Clinical findings • Examiner & Subject blinded to previous study findings • Examiner & Subject blinded to Std. reference results		
6	DATA ANALYSIS • Statistical Analysis Used	15%	15
7	RESULTS • Results Completely Displayed or Described • P-Value Displayed or Described • Confidence Interval Displayed or Described •		

Table 3

Included studies: Examiner / Subject / design / blinding

Author (year)	Examiner (number)	Study Subject	Study Design	Examiner Blinding
Harvey D (1991)	D.C. (n=27)	Mechanical Model	Cross-sectional	Blinded to fixation level and each other's findings
Moruzzi S (1993)	D.C. (n=50)	Mechanical Model	Cross-sectional	Blinded to fixation level and each other's findings
Jensen K (1993)	D.C. (n=45)	Mechanical Model	Cross-sectional	Blinded to fixation level and each other's findings
Sandmark H (1995)	P.T. (n= 1)	75 randomly selected males with acute neck pain (<= 1 wk)	Cross-sectional	Blinded to clinical presentation
Kristiansson, P (1996)	Not described	200 pregnant women with back pain	Cohort	Not described

Table 4

Average Quality Scores given in each of the 7 major criteria and the total and relative scores for each included article.

Author/Date	Study Subjects (Total 15)	Examiners (Total 10)	Reference Standard (Total 15)	Palpatory Test (Total 15)	Study Conditions (Total 15)	Data Analysis (Total 15)	Results (Total 15)	Total Mean Score	Relative Mean Score
Harvey, D / 1991	0	10	9	9	12	15	14.5	69.5	81.7
Moruzzi, S / 1993	0	10	12	15	15	15	15	82	96.5
Jensen, K / 1993	0	10	15	7	15	15	10	72	84.5
Sandmark, H / 1995	9.5	0	3.5	8	2	15	7.5	45.5	45.5
Kristiansson, P / 1996	8.5	0	9	11	2	15	10	55.5	55.5

Total Mean Score = Average of total absolute score obtained by each study

Relative Score = Total Mean score adjusted to 100% (to reflect “0” score given for subjects when mechanical models were used).

Table 5

Statistical analysis for Motion Palpation Studies using students and experienced practitioners

Examiners	Test	Harvey	Jensen
Both Groups	**PPV**	0.431	0.459
	NPV	0.9	0.902
	+LR	3.893	3.49
	-LR	0.564	0.403
Studen 0.54080 Tf 12 0 0 12 72 550.5599 Tm	**PPV**	0.437	0.367
	NPV	0.898	0.936
	+LR	3.71	4.23

Table 6

Spinal focus of the study, Reference standard used, Primary outcome, statistics, and author's conclusion.

Author (Year)	Spinal Focus	Reference Standard	Primary Outcome	Statistics	Author's Conclusion
Harvey D (1991)	Lumbar spine	Mechanical Model	Detect presence or absence of lumbar spine intersegmental motion restriction	Sensitivity Intern: 53.8%; Practitioner: 47.8%; Specificity Intern: 85.5%; Practitioner: 88% (PPV Pract. 42.3%, Interns 43.7%; NPV Pract. 90.3%, Interns 89.8%; +LR Pract. 4.05, Interns 3.7; -LR Pract. 0.592; Interns 0.54)	Intersegmental motion restriction palpation is more specific than sensitive
Moruzzi S (1993)	Lumbar spine	Mechanical Model	Detect accuracy of two types of spinal motion palpation procedures in correctly determining fixation	Sensitivity Lateral Flexion: 41.2%; PA springing: 42.8%; Specificity LF: 61.5%; PAS: 62.2% (PPV Post-Ant 28.6%; Lat. Flex. 30.6%; NPV Post-Ant 73.7%, Lat. Flex. 73.7%)	The palpation procedures as performed were not valid tests.
Jensen K (1993)	Lumbar spine	Mechanical Model	Detect presence or absence of	Sensitivity Interns: 72%; Practitioners: 52.6%;	Motion palpation is an accurate

			single and multiple intersegmental motion restrictions	Specificity Interns: 83.2%; Practitioners 78.6% (PPV Interns 46.2%; Pract. 45.5%; NPV Interns 93.7%; Pract 83%)	method for determining non-fixated segments but not accurate for determining fixated segments.
Sandmark H (1995)	Cervical spine	Pain reported by subjects	Assess presence or absence of pain upon palpation of facet joint	Sensitivity 82%; Specificity 79%; Positive Predictive Value=62%; NPV=91%	Palpation over the facet joint had better sensitivity and specificity than motion tests in study.
Kristianss on P (1996)	Lumbar spine	Visual Analog Scale	Assess the relationship between clinical back status and reported pain locations during and after pregnancy.	Thoracic DP Tenderness: Sensitivity 17.8%, Specificity 98.5%, Positive Predictive Value 72.2%, Negative Predictive Value 84.44%;; Lumbar DP Tenderness: Sens. 21.2%, Spec. 96.19%, PPV 61.76%, NPV 80.83%;;; Lumbar Percussion: Sens 5.1%, Spec. 100%, PPV 100%, NPV 78.44%.	Pain provocation tests were better at discriminating LBP than tests of configuration or mobility

DP = Digital Pressure

Pract. = Practitioners

+LR = positive Likelihood ratio

-LR = negative Likelihood ratio

Sens. = Sensitivity PPV = positive predictive value

NPV = negative predictive value

Appendix I

Inclusion and Exclusion Criteria for Validity Studies on Spinal Palpation as Applied to Title and Abstract Review

INCLUSION CRITERIA	EXCLUSION CRITERIA
The document must pertain to manual spinal (Cervical; Thoracic; Lumbar & surrounding para-spinal soft tissue, but not Sacrum or Pelvis) palpation procedures.	The data pertains to non-manual procedure(s).
The document includes measurement of validity or accuracy of spinal palpation. (Validity is the capability of the manual spinal palpation procedure to do what it is supposed to do; Accuracy is a measure of how well it actually does that!)	The document included a whole regimen of tests or methods; without separate data for each test, and/or the data for spinal palpatory procedure could not be retrieved.
The document must be a primary research study published in a peer reviewed journal, or dissertation. The document can be written in any language.	Although the document retrieved was relevant to the subject matter, it is anecdotal, speculative, or editorial in nature.
The primary research or monograph must have been published or accepted for publication. All	The document retrieved was inconsistent with the inclusion criteria.

Accuracy, accurate	Para-spinal	Manipulation, orthopedic [mh][sh][de]	Joint Instability [mh][sh][de]
Prediction, predict	Neck [mh][sh][de]	Manipulation, osteopathic [mh][sh][de]	Manipulable lesion
Predictive Value		Manipulation, spinal [mh][de]	Range of Motion, articular [mh][de] / Range of motion [sh]
Predictive value of tests [mh][sh][de]			Quality of Motion
Likelihood ratio			Tissue texture
Likelihood functions [mh][de]			Muscle tension / Muscle contraction [mh][sh][de]
ROC			Mobility
Discriminant validity [de]			Stiffness
Discriminant analysis [mh][de]			Myofascial pain syndromes [mh][sh][de]
Gold Standard			Apophyseal

Predictability			
Construct			
Criterion / Criterion related validity [sh]			#5: Validity Findings
Judgment [sh]			Measurement
Representation			Skin resistance / galvanic skin response [mh][fr]
Stability			Physiologic parameters
Generaliza*			Thermography [mh][sh][de]
Content / Content validity [sh]			Temperature
			Pain measurement [mh][sh][de]
			Radiography [mh] [de] [sh] X-rays [mh] [sh], x-ray [de]

[mh] MeSH -- Medicine Medical Subject Headings and is used for MEDLINE search

[de] descriptor - used for MANTIS and Biosis descriptor search

[sh] Subject headings - used for CINAHL Subject Headings search